

DTW 距離を用いた IC カード利用者のトリップパターンの類似度算出に関する基礎的検討

長岡技術科学大学大学院 環境システム工学専攻 学会員 ○新井佑枝

長岡技術科学大学 環境・建設系 正会員 西内裕晶

長岡技術科学大学 環境・建設系 正会員 佐野可寸志

1. はじめに

現在、交通に関する実態調査の最も基本的な調査の1つとして PT 調査がある。PT 調査は、人の動きを調べるために利用されていて、個人のある1日の動きを把握するものである。この1日の動きだけで評価するのではなく、複数の日数の動きがわかると、さらに詳細に個人がどのように行動しているか把握できると考えられる。詳細に個人の行動を捉え、日々の行動と違う行動を見つけることは、今後の公共交通活性化に向けての公共交通の料金や運行を検討する上での指標となりうると考えられる。

マら¹⁾は、中国の IC カードデータの時間・空間パターンを基に K 平均法およびラフ集合理論を使用してトリップパターンの規則性を分類している。西内ら²⁾の研究では、IC カードから得られる時系列データを用いて、利用者の時間的、空間的トリップパターンを定義し、日々の利用時刻と経路にパターンが存在しているのか、いないのかを分析した。その結果、公共交通の利用者の行動は、日々異なっているとの知見がある。しかし、これにはある日のトリップパターンが、完全に一致している場合のみで考えているため、本当に日々異なった行動を行っているのかについては疑問が残る。

そこで本研究では、ある日のトリップパターンが、完全に一致している場合だけでなく、実際には似ている場合を考慮し、各利用者がどのような特性を持っているか考察した。トリップパターンが、実際には似ている場合を判断する方法として、動的タイムワーピング(DTW : Dynamic Time Warping)距離を用いることとした。DTW 距離を用いて各利用者の代表的トリップパターンと他のトリップデータの類似性を評価し、日々の利用時刻と経路にパターンが存在しているか、いないのかを考察することを目的としている。

2. IC カードデータを用いた既存の研究

西内ら²⁾の研究では、各利用者が期間中に公共交通を同じ時間帯に利用しているか否かを判断するために、利用者のある1日において出現した乗車時刻、降車時刻の組み合わせ(60 分区分切り)を時間的トリップパターンとして定義する。例としてある人のトリップパターンが期間中に5日発生したとする(表-1 参照)。この場合、乗車時刻が 06:00~07:00、降車時刻が 07:00~08:00 の時間帯に1日目と2日目、4日目が属していることがわかる。3日目は時間帯とトリップ数が異なり、5日目は時間帯が異なることが見て取れるため、この利用者は5日行動している日数のうち、3日が全く同じ時間帯に行動していることになる。ここで移動日数のうち、どの程度同じ時間に行動をしているか考えるために、全日数の中で、同じ時間帯に行動している日数の割合を時間的トリップパターン依存度として定義している。表-1の例であれば期間中に移動した日数が5日で同じ時間帯に行動している日数が3日であるため、時間的トリップパターン依存度は0.6となる。

各利用者が期間中に公共交通を同じ場所で利用しているか、いないのかを判断するために、利用者のある期間中に移動した日数の中で同じ場所に移動している1日において出現した乗車場所、降車場所の組み合わせを空間的トリップパターンとして定義する。例としてある人のトリップパターンが期間中に5日発生したとする(表-2 参照)。この場合であれば1日目と4日目、5日目が A から B へと移動していることがわかる。こ

表-1 時間的トリップパターン

	乗車時刻	降車時刻	乗車時刻	降車時刻
1 日目	6:55	7:12		
2 日目	6:52	7:09		
3 日目	7:24	7:45	16:30	16:55
4 日目	6:45	7:01		
5 日目	6:03	6:15		

キーワード IC カードデータ, DTW 距離, 類似度, 空間的・時間的トリップパターン

連絡先 新潟県長岡市上富岡町 1-65-1 パストラルハイツ木村 202 TEL. 080-1225-5462 E-mail : s113306@stn.nagaokaut.ac.jp

表-2 空間的トリップパターン

	乗車場所	降車場所	乗車場所	降車場所
1日目	A	B		
2日目	B	A	A	B
3日目	A	C		
4日目	A	B		
5日目	A	B		

こで時間的トリップパターンと同じように移動日数のうち、どの程度同じ行動をしているか、空間的トリップパターン依存度として定義する。表-2の例であれば期間中に移動した日数が5日で同じ場所に移動している日数が3日であるため、時間的トリップパターン依存度は0.6となる。しかしBとCの場所がとても近かった場合においては、同じものとして判断すべきであると考えられる。そこで本研究では、西内ら²⁾が定義した空間的トリップパターンについて、DTW 距離を用いて類似したトリップパターンを考慮し、西内らの計算結果とどの程度異なるのかを確認する。また、公共交通利用者がどの程度同じ時間に移動しているか考察するために、利用者のある1日において出現した乗車時刻、降車時刻の組み合わせ(30分区分切り)を時間的トリップパターンとして定義することとした。

3. IC カードデータの概要

高知県のICカード「ですか」は、平成21年1月25日に導入され、(社)高知県バス協会・土佐電気鉄道(株)・土佐電ドリームサービス(株)・(株)県交北部交通・高知東部交通(株)で利用されている。³⁾

「ですか」の種類として、大人記名、大人無記名、身障者、小児、ナイスエイジの5種類に分類される。

本研究では、ICカード「ですか」の利用履歴を用いて、分析を行い、使用したデータの内容を表-3に示す。

4. 動的タイムワーピング距離(DTW 距離)

1) DTW 距離の計算方法

類似度を計算する方法には、さまざまな方法がある。

表-3 IC カード「ですか」の内容

項目	内容
データ日数	29日(6月4日は欠損) 平成22年6月1日～平成22年6月30日
利用者数	31,788人
データ内容	利用年月日、利用時刻、カードNo、券種種別、利用機関、乗車時刻、乗車停留所、降車停留所名、定期の有無、乗車距離、降車時刻

代表的なものとしては、ユークリッド距離などがあげられる。しかし、このユークリッド距離などで類似度を算出すると観測の間隔が違う場合などでは良い結果が得られない場合がある。

今回使用するデータは、時系列データであり、時間の間隔が同一ではないため、類似度を計算するには間隔のズレを考慮しなければいけない。そして、この時間のズレを考慮した距離として求めることができるのがDTW 距離である。連続データであるシーケンス間の距離を最小化するように時間軸方向にシーケンス長を調整する性質を持っているため、シーケンス長やサンプリングレートが異なるシーケンス間の類似度を計算することができる。⁴⁾そのため、時系列データも長さが違う場合でも時系列データ同士の比較を行うことができる。長さ m の時系列データ $x=(x_1, x_2, \dots, x_m)$ と長さ n の時系列データ $y=(y_1, y_2, \dots, y_n)$ 間のDTW 距離 $D(x, y)$ は下式で求めることができる。

$$D(x_i, y_j) = |x_i - y_j| + \min \begin{cases} D(x_{i-1}, y_j) \\ D(x_i, y_{j-1}) \\ D(x_{i-1}, y_{j-1}) \end{cases} \quad (1)$$

$(i=1, 2, \dots, m; j=1, 2, \dots, n)$

$$D(x_1, y_1) = |x_1 - y_1| \quad (2)$$

ここで x_i は x の第 i 要素から始まる接尾辞となる部分時系列データである。⁵⁾

次にDTW 距離の計算例を示す。2つの時系列データがあり、それを $x=(2, 4, 3, 2)$, $y=(3, 4, 3)$ とする。横軸を x 、縦軸を y の 4×3 の表として考える(図-1参照)。1段階目で一番左下のマス(x_1, y_1)を(2)式から計算する。2段階目では、縦軸と横軸の1列目を $D(x_1, y_j) = D(x_1, y_{j-1}) + |x_1 - y_j|$ と $D(x_i, y_1) = D(x_{i-1}, y_1) + |x_i - y_1|$ の計算により求める。3段階目では、左、下、左下の値が計算されているものを用いて(1)式により求める。

3段階までの計算を終えるとすべての要素が求まる。この一番後に計算した値、つまり $D(x_n, y_m)$ の値がDTW 距離となる。この値は、比較元のデータと似て

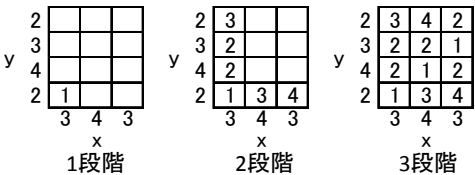


図-1 DTW 距離の求め方

いるものほど $D(x_n, y_m)$ の値が小さくなり、全く同じ時系列データ同士の場合は $D(x_n, y_m)$ の値は0になる。

2) 閾値の設定

DTW 距離は、算出された値が小さい程、比較したデータの類似度が高いと判断される。しかしながら、どの値が類似度が高いと判断できるかの閾値は、分析者が判断する必要がある。本稿では、空間的なトリップパターンに対して、どんな種類のトリップパターンが類似度が高いトリップパターンとして判断されたか、例として示す。図-2, 3 では太い赤線が比較される代表となり(以下、代表トリップ), DTW 距離の計算過程は、代表トリップと各色が似ているかどうかを比較している。具体的に、図-2 は、代表トリップに対する各色のトリップパターンの DTW 距離が 0.01 以上であり、計算上、類似していないと判断されたトリップパターンである。図-3 は、DTW 距離が 0.01 以下であり、それぞれが類似したトリップパターンであると計算された結果である。

各図を見てみると、明らかに違うようなトリップパターンのものが確実に図-2 に分類されていて、図-3 には明らかに違うトリップパターンは含まれていないと視覚的に確認できるため、本稿では、代表トリップパターンに対して他のトリップパターンが似ているか似ていないかの閾値を 0.01 と設定した。

5. トリップパターンの計算結果

1) 時間的・空間的トリップパターン依存度の関係について

本稿では、時間的トリップパターン依存度を公共交通を利用した時間帯が類似している組み合わせで最も多く出現した日数が、公共交通を利用した全トリップ日数に対する構成比と定義し、空間的トリップパターン依存度は、利用者の乗った場所と降りた場所の組み合わせが類似しているものの中で最も多く出現した日数が、全トリップ日数に対する構成比とした。2 つの値の関係を図-4, 5 に示す。図-4 は、DTW 距離を考慮していないもの、図-5 は、DTW 距離を考慮したものである。図中の点の大きさは、大きければ大きいほど、同じ空間的・時間的トリップパターン依存度の組合せである人の人数が多いことを示している。



図-2 似ていないと判断されたトリップパターン



図-3 似ていると判断されたトリップパターン

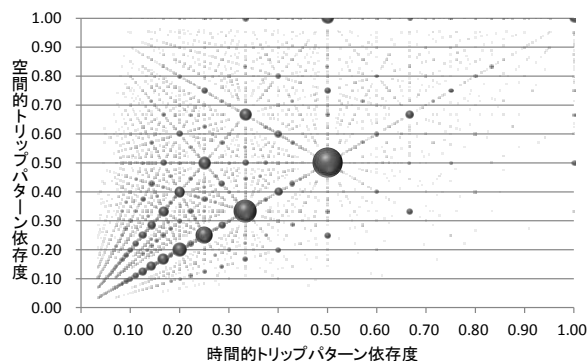


図-4 時間的・空間的トリップパターンの関係
(DTW 距離考慮せず)

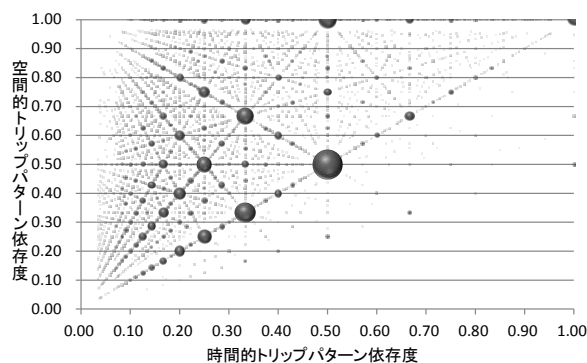


図-5 時間的・空間的トリップパターンの関係
(DTW 距離考慮)

図-4はDTW距離が考慮されていないため、完璧に一致したトリップパターンで無いと同じトリップパターンとして判断されないで、空間的トリップパターン依存度の値が全体的に小さい値になる(縦軸の値が下方に多く存在する)傾向が見られる。一方で、図-5はDTW距離によりトリップパターンの類似度を考慮したものであり、空間的トリップパターン依存度が図-4に比べて比較的高い傾向にある。つまり、毎日同じ行動をしていると判断されたトリップパターンが増加した結果である。このことから、DTW距離を用いることで空間的なトリップパターンの類似度を考慮できたと分かる。

2) 空間的トリップパターン依存度の比較

DTW距離を考慮したものと考慮していないもので、空間的トリップパターン依存度の値を比較したものを図-6に示す。

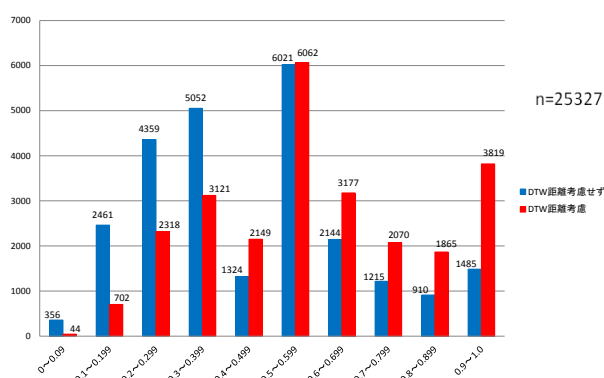


図-6 空間的トリップパターン依存度の比較

この結果を見ると、DTW距離を考慮しないものは、0.5以下の値が多いため、多数の人が、出現日数の半数以上は、毎日異なる経路で移動している結果となっているのに対し、DTW距離を考慮することで、0.5以上の値が増えている。つまり、出現日数の半数以上は、似ている経路で移動しているパターンが存在していることを示すことができた。

6. おわりに

本研究では、ICカード「ですか」を用いて、ICカード利用者の空間的・時間的トリップパターンを定義した。DTW距離を使用して空間的トリップパターン依存度を考えることで類似しているトリップパターンを

考慮して分析する可能性を示すことができた。すなわち、以前の研究¹⁾で、毎日違う行動をしていると判断されたトリップパターンであってもDTW距離を考慮することで類似したトリップパターンを考慮し、より現実的な日々の行動パターンを解析する可能性を示すことができたと言える。特に本稿では、DTW距離により類似していると判断した閾値は0.01であった。しかし、分析対象によって、より厳密に類似度を考慮する場合は、閾値をより低い値に設定するなどの検討をする必要がある。

今後の課題としては、時間的トリップパターンの考え方について考慮する必要がある。本稿では、30分区分りで分けて考えていたが、類似しているものでも違う行動と判断されてしまう場合がある。例えば、07:00と07:05であれば一緒に判断されるが、06:55と07:00は全く違うものと判断されてしまうため、DTW距離の適用も含めて再検討する必要がある。また、結果より日々似ている行動をしている人が多いと分かったが、そのような人の似ていないと判断されたトリップパターンなどについても、似ていないトリップパターンを行った理由などの考察をする必要があるだろう。更に、DTW距離の閾値と実際のトリップパターンの形状について、より詳細な評価・考察が必要であると考えられる。

<参考文献>

- 1) Ma X., Wu Y., Wang Y., Chen F. and Liu J. : Mining smart card data for transit riders' travel patterns, *Transportation Research Part C*, Vol.36, pp.1-12, 2013
- 2) 西内裕晶, 轟朝幸: 交通マーケティング手法検討のためのICカードデータを利用した利用者行動特性の把握, 土木学会論文集 F3(土木情報学), 68巻, 2号, II-8頁~II-17頁, 2013
- 3) “ですか”, 株式会社ですか <http://www.desuca.co.jp/>, 参照 2013.9.30
- 4) 豊田真智子, 櫻井保志, 市川俊一: データストリームのための部分シーケンスマッチング, DEIM Forum 2009, 2009
- 5) 石川雅弘, 吉川昂伯, 陳漢雄, 古瀬一隆, 大保信夫: 類似部分区間検索のためのタイムワーピング距離の下限值計算, DBSJ Letters Vol.6, No.1, 2007